09/384,922
Q55464
1/8/01

Europäisches Patentamt | European Patent Office | Office européen des brevets

OIPE JC89
SEP 30 1999
PATENT & TRADEMARK OFFICE

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

99402021.2

CERTIFIED COPY OF PRIORITY DOCUMENT

Der Präsident des Europäischen Patentamts; Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets p.o.

A. FOTTE

DEN HAAG, DEN
THE HAGUE, 01/09/99
LA HAYE, LE

EPA/EPO/OEB Form 1014 - 02.91

#2 11-10-99 GP2731

OIPE JC89 SEP 30 1999 PATENT & TRADEMARK OFFICE

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Peter Paul Camille DE SCHRIJVER, et al.

Attorney Docket Q55464

Appln. No.: 09/384,422

Group Art Unit: 2731

Filed: August 27, 1999

Examiner: Not yet assigned

For: METHOD FOR TRANSPORTING DATA, A RELATED DATA TRANSMITTING ELEMENT AND A DATA RECEIVING ELEMENT

RECEIVED OCT -4 1999 TC 2700 MAIL ROOM

SUBMISSION OF PRIORITY DOCUMENT

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Submitted herewith is a certified copy of the priority document on which a claim to priority was made under 35 U.S.C. § 119. The Examiner is respectfully requested to acknowledge receipt of said priority document.

Respectfully submitted,

David J. Cushing
Registration No. 28,703

SUGHRUE, MION, ZINN,
MACPEAK & SEAS, PLLC
2100 Pennsylvania Avenue, N.W.
Washington, D.C. 20037-3212
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

Enclosures: **CERTIFIED COPY OF EUROPEAN PATENT APPLICATION NO. 99402021.2**

Date: SEP 30 1999

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 99402021.2
Demande n°:

Anmeldetag:
Date of filing: 09/08/99
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
ALCATEL
75008 Paris
FRANCE

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Method for transporting data, a related data transmitting element and a data receiving element

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

METHOD FOR TRANSPORTING DATA, A RELATED DATA TRANSMITTING ELEMENT AND A DATA RECEIVING ELEMENT

The present invention relates to a method as described in preamble of claim1 and claim2, to a related data transmitting device as described in preamble of claim 3, to a related data receiving element as described in preamble of claim 5 and claim 6 and the related software modules as described in the preamble of claims 7, 9 and 10.

Such a method and related network elements are generally known in the art. A data transmitting network element sends data-packets towards a destination through a communications network via an edge network element of the communications network. In order to prevent from network congestion caused by a data transmitting network elements transmitting data without taking capacity-constraints and other possible constraints into account, a maximum service level is negotiated between each of such data transmitting elements and the edge element. Such a service level specification nowadays is negotiated between both parties in a static way, for example by phone call or by e-mail. There may be a policing means within the edge element to check if incoming data-packet flows are in conformance with the agreed service level specification. In case that the data transmitting network element at a certain moment of time needs another service level, this should be renegotiated in the same static way. Then the service level providing system should be updated to actually provide the agreed service level. This is a very time-consuming and inefficient way of negotiating service level specifications. Currently there is no automatic negotiation and re-negotiation and the subsequent updating of the Service Level Specification possible.

An object of the present invention is to provide a method of the above known type but wherein the service level specification negotiation and at the same time the adaptation is performed in a more efficient way.

According to the present invention, this object is achieved by the method as claimed in claims 1 and 2, the related elements as claimed in claim 3 and claims 5 and 6 and the related software modules as claimed in claims 7, 9 and 10.

Indeed, by negotiating a service level specification between a data transmitter and a data receiver and vice versa on both ends of a Point-to-Point connection, based on predetermined criteria, during the establishment of the Point-to-Point connection by sending Internet Protocol Control Protocol messages, a service level specification is agreed between both parties and this service level specification is updated at the data receiving element. For this purpose there are new options defined for the Internet Protocol Control Protocol that forms part of the Point-to-Point Protocol, where this options contain service level specific parameters.

This negotiation can be initiated by the data transmitting element as in claims 1, 5 and 7, or by the data receiving element as in claims 2, 6 and 10. In the latter case this is done by sending Internet Protocol Control Protocol messages to the data transmitting element if there are conditions that require the data receiving element to adapt the provided service level.

A further characteristic feature of the present invention is described in claims 4 and 8. The received propose for a service level at the data transmitting element may be not satisfying in one or another way for the data transmitting element. If so the data transmitting element is able to notice this and to formulate another request for a service level that is sufficient for the data transmitting element. Re-negotiation of an existing service level specification is also possible by sending Internet Protocol Control Protocol messages requesting another service level specification.

The above and other objects and features of the invention will become more apparent and the invention itself will be best understood by referring to the following description of an embodiment taken in conjunction with the accompanying drawings wherein:

FIG. 1 represents an internet network INNW wherein the method for transporting data is performed; and FIG.2 represents the functional built-up of the data transmitting element DTE and the data receiving element DRE as presented in FIG. 1.

In the following paragraphs, referring to the drawings, an implementation of the method according to the present invention will be described. Special attention will be drawn to the components of the data transmitting element and the components of the data receiving element as presented in FIG. 2. In the second paragraph, all connections between the before mentioned network elements and described components are defined. In the succeeding paragraph the actual execution of the service level specification negotiation is described.

The essential elements of this embodiment of the present invention are an internet network INNW, a data transmitting element DTE and a data receiving element DRE. In this embodiment the data transmitting element DTE is a customer premises equipment element, a personal computer.

This personal computer is used for sending data towards the data receiving network element DRE. The sending is done according to a pre-agreed service level specification defining quality of service aspects for the user of the personal computer. For example, in the service level specification specified Quality of service parameter is the maximum bit-rate of a certain diffserv class the sender is allowed to send .

In order to keep simplicity in this description it is chosen to only describe one personal computer DTE connected to the network, although there normally will be a plurality of such user-terminals.

Further there is a data receiving element DRE, that in this embodiment is chosen to be a network access server situated at the edge of the internet network INNW. This network access server provides the personal computer DTE access to the internet network INNW. Besides this, the network access server DRE

may take care of policing the data the personal computer DTE sends towards the network access server DRE.

The personal computer DTE is to be connected to the internet via a Point-to Point Protocol connection, further referred to as a PPP-connection.

The data transmitting element DTE as presented in FIG. 2 is built up of a data sending means DSM that takes care of sending data towards the data receiving element and a service level requesting means SL_R_M that is able to send a request to the data receiving element DRE for a predetermined service level for sending data, using an Internet Protocol Control Protocol request message for this purpose. Further there is a service level propose receiving means SLP_R_M that is able to receive an Internet Protocol Control Protocol propose for the service level to be provided and subsequently to notify the data sending means DSM of the propose for the service level to be provided. The data transmitting element DTE also comprises a service level propose renegotiating means SLP_RN_M that is able to check if a received Internet Protocol Control Protocol propose for a service level is satisfying and if not to formulate another request for a service level.

The service level requesting means SL_R_M has an output-terminal that is at the same time an output-terminal O_1 of the data transmitting element DTE. The data sending means DSM has an output-terminal that is at the same time an output-terminal O_2 of the data transmitting element DTE.

Further, the service level propose receiving means SLP_R_M, has an input-terminal that is at the same time an input-terminal I_1 of the data transmitting element DTE and besides this an output-terminal that is coupled to an input-terminal of the data sending means DSM. The service level propose renegotiating means SLP_RN_M is coupled between a second output-terminal of the service level propose receiving means SLP_R_M and an input-terminal of the service level requesting means SL_R_M.

The data receiving element DRE as presented in FIG. 2 comprises a receiving means DRM that is able to receive data from the data transmitting

element DTE. Further the data receiving element DRE contains service level request reception means SLR_Re_M that is adapted to receive a service level specification request from the data transmitting element DTE using an Internet Protocol Control Protocol message and a service level negotiating and proposing means SL_NP_M that is able to determine a service level specification based on at least one predetermined criterion and to formulate a propose for the service level specification. Then there is a service level proposal sending means SLP_S_M, that is adapted to send the propose for the service level specification using an Internet Protocol Control Protocol message.

The service level request reception means SLR_Re_M has an input-terminal that is at the same time an input-terminal I_2 of the data receiving element DRE and an output-terminal that is coupled to an input-terminal of the service level negotiating and proposing means SL_NP_M that in its turn is coupled with an output-terminal to an input-terminal of the service level request reception means SLR_Re_M. The service level proposal sending means SLP_S_M has an output-terminal that is at he same time an output-terminal O_3 of the data receiving element DRE. Then the data receiving means DRM contains an input-terminal that is at the same time an input-terminal I_3 of the data receiving element DRE.

In order to explain the operation of the present invention it is assumed that the personal computer DTE needs a specific quality of service that is specified in a service level specification. It is assumed that the personal computer DTE needs to establish a dial-in connection using the Point-to-Point Protocol. In a phase of this connection establishment, Internet Protocol Control messages, further referred to as IPCP-message are sent towards the network access server DRE to negotiate several connection–parameters, for example an Internet Protocol-address.

It is further to be noticed that all in the following mentioned IPCP-messages or requests are not the normally used messages or requests but a

modified message or request in that sense that each IPCP-message may contain additional options carrying service level specification negotiable parameters.

In order to request a specific service level specification, the service level requesting means SL_R_M of the personal computer DTE sends an Internet Protocol Control Protocol request towards the network access server DRE for assigning another service level for sending data. The service level request reception means SLR_Re_M in its turn receives the Internet Protocol Control Protocol request for the service level from the data transmitting element DTE and forwards the request to the service level negotiating and proposing means SL_NP_M. The service level negotiating and proposing means SL_NP_M then determines a service level based on at least one predetermined criterion and subsequently formulates a propose for the service level to be provided. Such a criterion could be the current traffic on the outgoing links of the data receiving element DTE or a criterion available from a global network element like a policy server that can manage the entire internet network INNW.

Subsequently the service level negotiating and proposing means SL_NP_M hands the propose over to the service level proposal sending means SLP_R_M that in its turn sends an Internet Protocol Control Protocol message that contains the relevant parameters of the propose for the service level to the service level propose receiving means SLP_R_M of the data transmitting element DTE which receives the Internet Protocol Control Protocol propose for the service level. The service level propose receiving means SLP_R_M notifies the data sending means DSM of the propose for the service level.

The service level propose renegotiating means SLP_RN_M receives the forwarded service level propose from the service level propose receiving means SLP_R_M and subsequently checks if this Internet Protocol Control Protocol propose for the service level is satisfying. If not satisfying the service level propose renegotiating means SLP_RN_M formulates another request for the desired service level and forwards this towards the service level requesting means SL_R_M that in its turn further handles the request.

There may be a policing means (not shown in any of the figures) available within the data receiving element to determine if the data transmitting element sends data in compliance with the agreed service level.

It is to be remarked that the data receiving element DRE is also able to initiate a service level specification negotiation or re-negotiation by sending an unsolicited propose.

It is further to be remarked that the data transmitting element DTE and the data receiving element DRE in another embodiment may be implemented by two coupled routers.

Although the above embodiment of the invention has been described by means of functional blocks, their detailed realisation based on this functional description should be obvious for a person skilled in the art and is therefore not described.

While the principles of the invention have been described above in connection with specific apparatus, it is to be clearly understood that this description is made only by way of example and not as a limitation on the scope of the invention, as defined in the appended claims.

CLAIMS

1. Method for transporting data between a data transmitting network element (DTE) and a data receiving network element (DRE) via a point-to-point connection in a communications network comprising at least one of said data transmitting network element (DTE) and at least one of said data receiving network element (DRE), said method comprising the steps of:

a. sending data towards said data receiving element (DRE) by said data transmitting element (DTE) via said point-to-point connection;

b. receiving said data sent by said data transmitting element (DTE) by said data receiving element (DRE), **CHARACTERISED IN THAT SAID** method further comprises the following steps:

c. sending an Internet Protocol Control Protocol request for a service level of sending data to said data receiving element (DRE) by said data transmitting element (DTE);

d. receiving said Internet Protocol Control Protocol service level request from said data transmitting element (DTE) by said data receiving element (DRE);

e. determining by said data receiving element (DRE) a service level based on at least one predetermined criterion and formulating an Internet Protocol Control Protocol propose of said service level that can be provided to said data sending element (DSE);

f. sending said Internet Protocol Control Protocol propose of said service level towards said data transmitting element (DTE);

g. receiving said Internet Protocol Control Protocol propose of said service level and using said propose of said service level by said data transmitting element (DTE).

2. Method for transporting data between a data transmitting network element (DTE) and a data receiving network element (DRE) via a point-to-point

connection in a communications network comprising at least one of said data transmitting network element (DTE) and at least one of said data receiving network element (DRE), said method comprising the steps of:

a. sending data towards said data receiving element (DRE) by said data transmitting element (DTE) via said point-to-point connection;

b. receiving said data sent by said data transmitting element (DTE) by said data receiving element (DRE), **CHARACTERISED IN THAT SAID** method further comprises the following steps:

c. determining by said data receiving element (DRE) a service level based on at least one predetermined criterion and formulating an Internet Protocol Control Protocol propose of said service level that can be provided to said data sending element (DSE);

d. sending said Internet Protocol Control Protocol propose of said service level towards said data transmitting element (DTE);

e. receiving said Internet Protocol Control Protocol propose of said service level and using said propose of said service level by said data transmitting element (DTE).

3. Data transmitting element (DTE), to be used for sending data, over a link through a communications network towards a data receiving element (DRE), said data transmitting element comprising the following means:

a. data sending means (DSM), adapted to send data towards said data receiving element, **CHARACTERISED IN THAT SAID** data transmitting element (DTE) further comprises the following means:

b. service level requesting means (SL_R_M), adapted to request said data receiving element (DRE) for a service level for sending said data using an Internet Protocol Control Protocol message;

c. service level propose receiving means (SLP_R_M), coupled with an output to an input of said data sending means (DSM) and adapted to receive an Internet Protocol Control Protocol propose for said service level and to notify said

data sending means (DSM) of said propose for said service level.

4. Data transmitting element (DTE) according to claim 3, **CHARACTERISED IN THAT SAID** data transmitting element (DTE), further comprises a service level propose renegotiating means (SLP_RN_M), coupled between an output-terminal of said service level propose receiving means (SLP_R_M) and an input-terminal of said service level requesting means (SL_R_M) and adapted to check if said Internet Protocol Control Protocol propose for said service level is satisfying and if not, to formulate another request for said service level.

5. Data receiving element (DRE), to be used for receiving data, over a link through a communications network from a data transmitting element (DTE), said data receiving element (DRE) comprising the following means:

a. data receiving means (DRM), adapted to receive data from said data transmitting element, **CHARACTERISED IN THAT SAID** data receiving element (DRE) further comprises the following means:

b. service level request reception means (SLR_Re_M), adapted to receive a service level request from said data transmitting element (DTE) using an Internet Protocol Control Protocol message;

c. service level negotiating and proposing means (SL_NP_M), coupled with an input to an output of said service level request reception means (SLR_Re_M) and adapted to determine a service level based on at least one predetermined criterion and to formulate a propose for said service level;

d. service level proposal sending means (SLP_S_M), coupled with an input to an output of said service level negotiating and proposing means (SL_NP_M) and adapted to send said propose for said service level using an Internet Protocol Control Protocol message.

6. Data receiving element (DRE), to be used for receiving data, over a link through a communications network from a data transmitting element (DTE), said data receiving element (DRE) comprising the following means:

a. data receiving means (DRM), adapted to receive data from said data transmitting element, **CHARACTERISED IN THAT SAID** data receiving element (DRE) further comprises the following means:

b. service level negotiating and proposing means (SL_NP_M), adapted to determine a service level based on at least one predetermined criterion and to formulate a propose for said service level;

c. service level proposal sending means (SLP_S_M), coupled with an input to an output of said service level negotiating and proposing means (SL_NP_M) and adapted to send said propose for said service level using an Internet Protocol Control Protocol message.

7. Software module for running on a processing system for inclusion in a data transmitting element (DTE), for sending data over a link through a communications network towards a data receiving element (DRE), said software module comprising the following sub-modules:

a. data sending sub-module, adapted to send data towards said data receiving element, **CHARACTERISED IN THAT SAID** software module further comprises the following sub-modules:

b. service level requesting sub-module, adapted to request said data receiving element (DRE) for a service level for sending said data using an Internet Protocol Control Protocol message;

c. service level propose receiving sub-module, adapted to receive an Internet Protocol Control Protocol propose for said service level and to notify said data sending sub-module of said propose for said service level.

8. Software module according to claim 7, **CHARACTERISED IN THAT SAID** software module, further comprises a service level propose

renegotiating sub-module, co-operating with said service level propose receiving sub-module and said service level requesting sub-module and adapted to check if said Internet Protocol Control Protocol propose for said service level is satisfying and if not, to formulate another request for said service level.

9. Software module for running on a processing system for inclusion in a data receiving element (DRE), for receiving data over a link through a communications network from a data transmitting element (DTE), said software module comprising the following sub-modules:

a. data receiving sub-module, adapted to receive data from said data transmitting element (DTE), **CHARACTERISED IN THAT SAID** software module further comprises the following sub-modules:

b. service level request reception sub-module, adapted to receive a service level request from said data transmitting element (DTE) using an Internet Protocol Control Protocol message;

c. service level negotiating and proposing sub-module, co-operating with said service level request reception sub-module and adapted to determine a service level based on at least one predetermined criterion and to formulate a propose for said service level;

d. service level proposal sending sub-module, co-operating with said service level negotiating and proposing sub-module and adapted to send said propose for said service level using an Internet Protocol Control Protocol message.

10. Software module for running on a processing system for inclusion in a data receiving element (DRE), for receiving data over a link through a communications network from a data transmitting element (DTE), said software module comprising the following sub-modules:

a. data receiving sub-module, adapted to receive data from said data transmitting element (DTE), **CHARACTERISED IN THAT SAID** software module further comprises the following means:

d. service level negotiating and proposing sub-module, adapted to determine a service level based on at least one predetermined criterion and to formulate a propose for said service level;

e. service level proposal sending sub-module, co-operating with said service level negotiating and proposing sub-module and adapted to send said propose for said service level using an Internet Protocol Control Protocol message.

ABSTRACT

METHOD FOR TRANSPORTING DATA, A RELATED DATA TRANSMITTING ELEMENT AND A DATA RECEIVING ELEMENT

The present invention relates to a method for transporting data between a data transmitting network element and a data receiving network element via a point-to-point connection in a communications network. This communications network comprises at least one transmitting network element and at least one data receiving network element.

An Internet Protocol Control Protocol request for a service level of sending data is sent to the data receiving element by the data transmitting element. This Internet Protocol Control Protocol request is received by the data receiving element. The data receiving element then determines a service level based on a predetermined criterion and formulates an Internet Protocol Control Protocol propose of the service level that can be provided to the data sending element. Subsequently the Internet Protocol Control Protocol propose of the service level is sent towards the data transmitting element. The Internet Protocol Control Protocol propose including the service level propose in its turn is received by the data transmitting element and used for sending data towards the data receiving element.

It is also possible to initiate the (re-)negotiation by the data receiving network element, by sending Internet Protocol Control Protocol messages to the data transmitting element.

As an alternative the data transmitting element is able to notice that a provided service level is not satisfying and that it formulates another request for a service level that is satisfying for the data transmitting element. Re-negotiation of an existing service level specification is also possible by sending Internet Protocol Control Protocol messages requesting another service level specification.

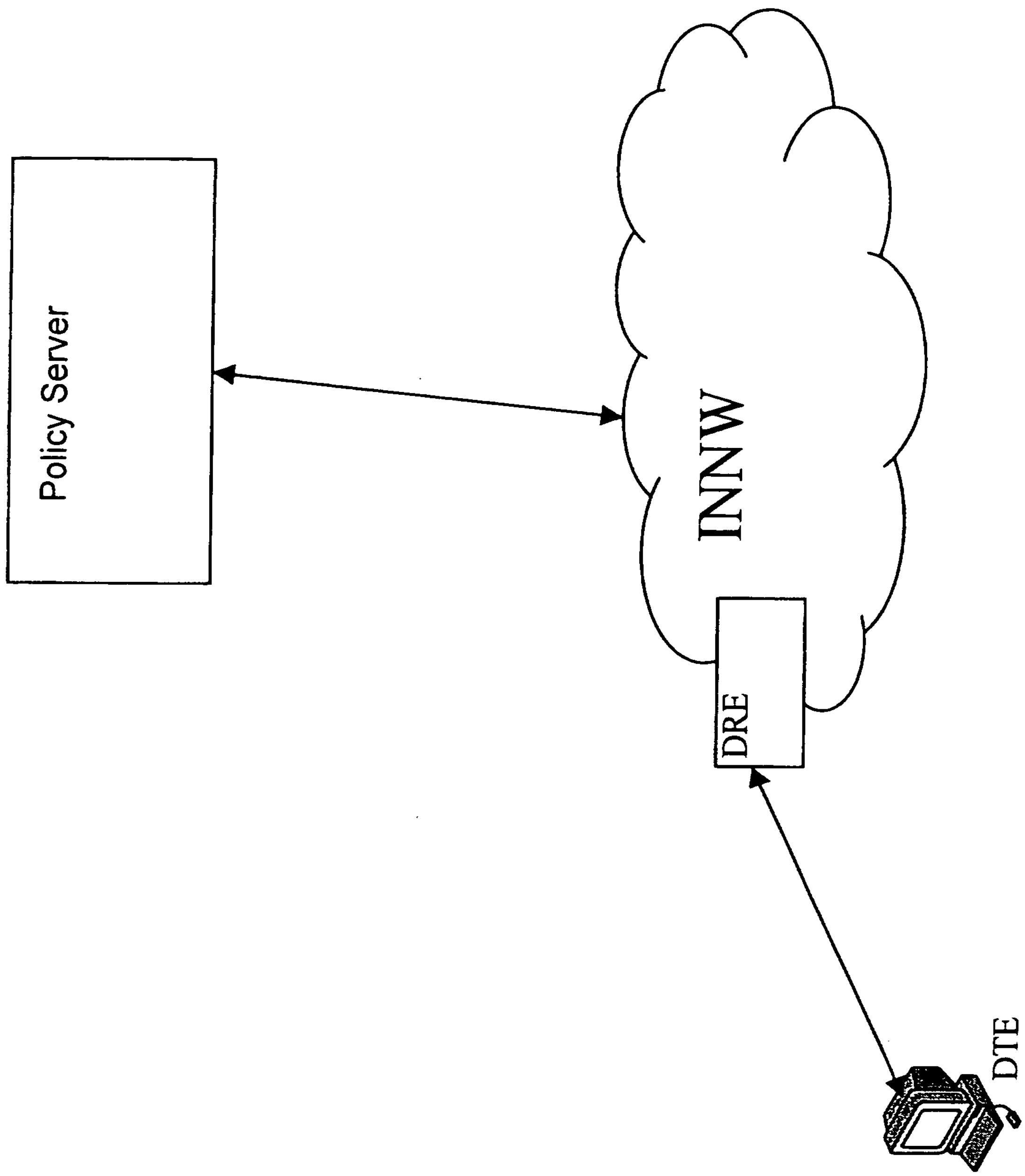

FIG. 1

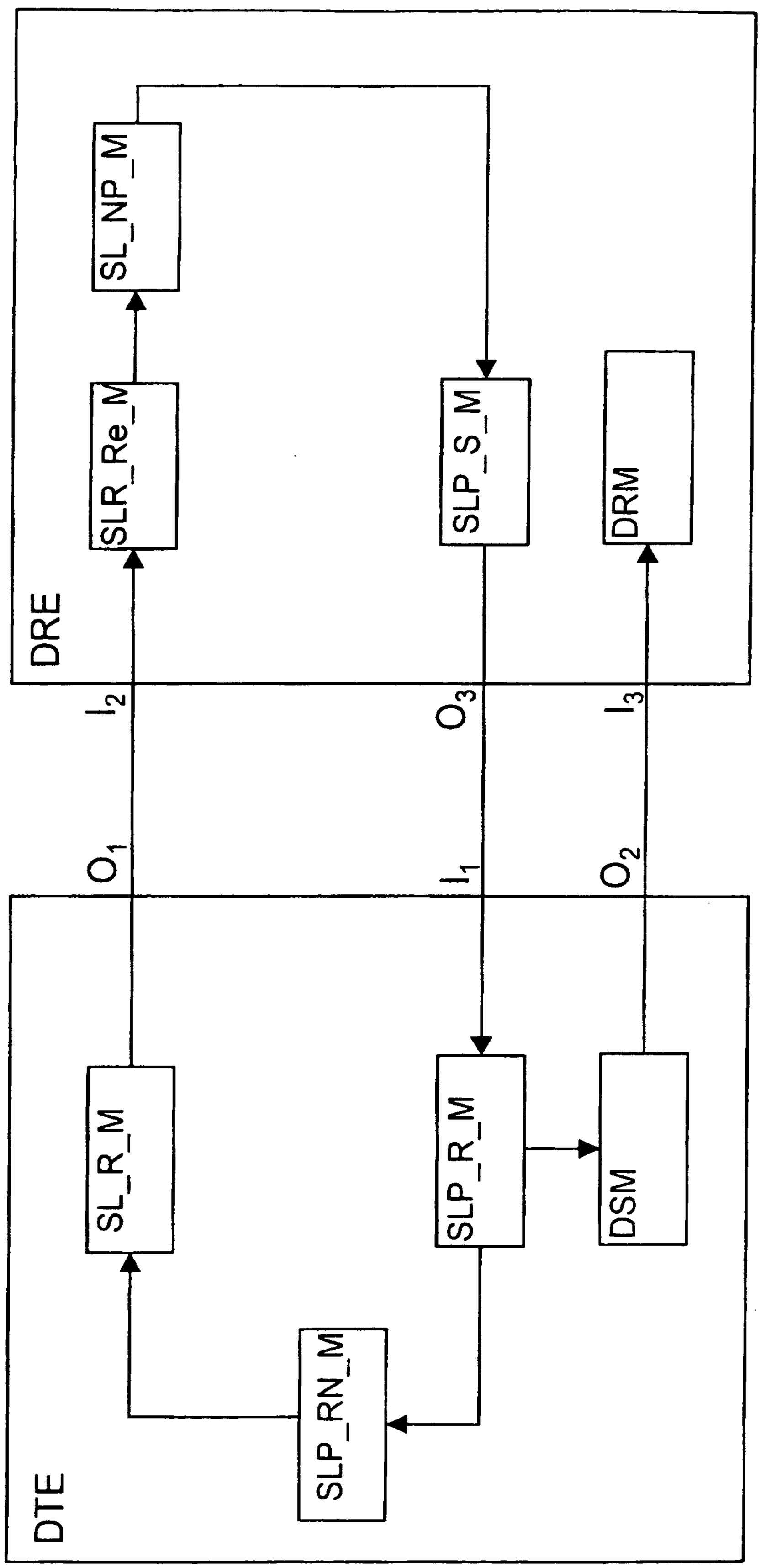
DRE
SL_NP_M
SLR_Re_M
SLP_S_M
DRM
I_2
O_3
I_3
O_1
I_1
O_2
DTE
SL_R_M
SLP_R_M
DSM
SLP_RN_M

FIG. 2

Potts

Name: **John Potts**
Address: **Rijswijk Main, Room 01.01, Tel: 3161**

Userid: JP02017

Printer: P0242
Requested: 01/09/99 16:22:15

JobNo: P30WS00
Batch: 1 of 1

Phoenix printout

99480017

2.00

Multiple ARP functionality for an IP data transmission system

Technical field

The present invention deals with a new way for load balancing outgoing IP packets from an IP host such as a large Web server, and relates in particular to a multiple ARP functionality for an IP data transmission system.

Background

Modern digital networks are made to operate over different transmission media and interconnect upon request a very large number of users (e.g. hosts) and applications through fairly complex digital communication networks.

Due to the large variety of users' profiles and distributed applications, the traffic is becoming more and more bandwidth consuming, non-deterministic and requiring more connectivity. This has been the driver for the emergence of fast packet switching techniques in which data from different origins are chopped into fixed or variable length packets or datagrams, and then transferred, over high speed digital networks, between a data source and a target terminal equipment.

Several types of networks have been installed throughout the world, which need to be interconnected (e.g. via so called Routers) to optimize the possibilities of organizing traffic between source hosts and target hosts located anywhere in the world. This is made possible by using so-called internet-working.

Internetwork (also referred to as internet) facilities use a set of networking protocols such as Transmission Control Protocol/Internet Protocol (TCP/IP) developed to allow cooperating host computers to share resources across the internet-work. TCP/IP is a set of data communication protocols that are

referred to as internet protocol (IP) suite. Because TCP and IP are the best known, it has become common to use the term TCP/IP to refer to the whole protocol family. TCP and IP are two of the protocols in this suite. Other protocols of the suite are User Datagram Protocol (UDP), Address Resolution Protocol (ARP), Real Time Protocol (RTP) etc…

An internet may thus be a collection of heterogeneous and independent networks using TCP/IP, and connected together by routers. The administrative responsibilities for an internet (e.g. to assign IP addresses and domain names) can be within a single network (LAN) or distributed among multiple networks.

When a communication of data has to be established from a source host to a particular IP destination over an IP network, there is a number of methods to determine the first hop router of the network towards this destination. These include running (or snooping) dynamic routing protocol such as Routing Information Protocol (RIP) or Open Shortest Path First (OSPF) version, running an ICMP router discovery client or using a statically configured default route.

Running a dynamic routing protocol on every end-host may be infeasible for a number of reasons, including administrative overhead, processing overhead, security issues, or lack of a protocol implementation for some platforms. Neighbor or router discovery protocols may require active participation by all hosts on a network, leading to large timer values to reduce protocol overhead in face of large numbers of hosts. This can result in a significant delay in the detection of a lost (i.e., dead) neighbor, which may introduce unacceptably long "black hole" periods.

The use of a statically configured default route is quite popular, it minimizes configuration and processing overhead on the end-host and is supported by virtually every IP implementation. This mode of operation is likely to persist as Dynamic Host Configuration Protocols (DHCP) are deployed, which typically provide configuration for an end-host IP address and

default gateway. However, this creates a single point of failure. Loss of the default router results in a catastrophic event, isolating all end-hosts that are unable to detect any alternate path that may be available.

One solution to solve this problem is to allow hosts to appear to use a single router and to maintain connectivity even if the actual first hop router they are using fails. Multiple routers participate in this protocol and in concert create the illusion of a single virtual router. The protocol ensures that one and only one of the routers is forwarding packets on behalf of the virtual router. End hosts forward their packets to the virtual router. The router forwarding packets is known as the active router. A standby router is selected to replace the active router should it fail. The protocol provides a mechanism for determining active and standby routers, using the IP addresses on the participating routers. If an active router fails, a standby router can take over without a major interruption in the host's connectivity.

Another similar approach is the use of Virtual Router Redundancy Protocol (VRRP) designed to eliminate the single point of failure inherent in the static default routed environment. VRRP specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers on a LAN. The VRRP router controlling the IP address(es) associated with a virtual router is called the Master, and forwards packets sent to these IP addresses. The election process provides dynamic fail-over in the forwarding responsibility should the Master become unavailable. Any of the virtual router's IP addresses on a LAN can then be used as the default first hop router by end-hosts. The advantage gained from using VRRP is a higher availability default path without requiring configuration of dynamic routing or router discovery protocols on every end-host.

Unfortunately the two above solutions cannot provide load balancing for a given host's traffic because only the router that answered the ARP is used. Also, customers are reluctant

to change their main router configuration to enable such a function.

Summary of the invention

Accordingly, the object of the invention is to provide a data transmission system including an IP network wherein it is the IP host which selects directly the default router thereby improving load balancing and high availability.

Another object of the invention is to enable an IP source host to be aware of the availability of a set of candidate default routers and to select one of them dynamically, ensuring both load balancing and high availability.

Another object of the invention is a method of selecting a router amongst a set of routers for an IP host in a data transmission system including an IP network.

Therefore, the invention relates to a data transmission system for transmitting packetized data from an IP host having at least an IP layer and a network layer to a plurality of work-stations by the intermediary of an IP network and wherein the IP host is connected to the IP network via a layer 2 network interfacing the IP network by a set of routers, the IP host further including a Multiple Address Resolution Protocol (MARP) layer between the IP layer and the network layer for selecting one of the set of routers in response to the next hop IP address provided by the IP layer to the multiple ARP layer when a packet of data is to be transmitted from the IP host to one of the workstations.

Brief description of the drawings

The above and other objects, features and advantages of the invention will be better understood by reading the following more particular description of the invention in conjunction with the accompanying drawings wherein :

Fig. 1 represents schematically a data transmission system wherein an IP host can select one router amongst a set of routers according to the invention.

Fig. 2A and 2B represent respectively the MARP table and the ARP table used in combination to achieve the method according to the invention.

Fig. 3 is a flow chart of the method of selecting a router according to the invention.

Detailed description of the invention

In reference to Fig. 1, the invention is implemented in a data transmission system wherein an IP host has to transmit data to one or several workstations 12, 14 via an IP network 16 such as Internet. It can be assumed that IP host 10 is connected to IP network 16 by means of a layer 2 network such as Local Area Network (LAN) 18 which is interfacing IP network 16 by a set of input routers 20, 22 and 24. The IP packets are routed over the IP network via a plurality of routers (not shown) until an output router 26 connected directly (or by means of a layer 2 network) to workstations 12 or 14.

As illustrated in Fig. 1, to communicate over the IP network, IP host 10 must implement a layered set of protocols 28 referred as the Internet protocol suite. Without the invention the protocol suite would be used as follows :

- the application layer 30 (level 5) generates a data stream to be sent and passes this data stream to a transport layer,
- the transport layer (level 4) such as TCP layer 32, segments the data stream into packets and passes the packet to the IP layer for routing to the destination IP address with an added TCP Header,
- the IP layer 34 finds the next hop IP address based upon the destination IP address. Normally, with the IP Host

which does not run a routing protocol, this address is a default entry that leads to a default router.

- IP layer 34 passes the IP packet to the network layer (not shown) with an added IP header information. As a side parameter, the IP layer informs the network layer of the next hop IP address.
- the network layer resolves the next hop IP address into a network address of the default router using the ARP protocol and transmits the packet over the IP network.

The invention introduces a new layer between IP layer 34 and the network layer, a Multiple ARP (MRAP) layer 36. Therefore, IP layer 34 passes the packet and the next hop IP address to MARP layer 36 instead of the network layer. As explained below, this MARP layer runs an algorithm to determine the best physical router 20, 22 or 24 based on parameters defined in the packet such as source and destination addresses and ports.

At the destination workstation 12, a reciprocal protocol suite 38 is implemented. Namely, the network layer passes the IP packets to IP layer 40 which transfers the packets to TCP layer 42 for reassembling them into a data stream communicated to the application layer 44. Note that workstation 12 does not include a MARP layer since such a layer is not required for receiving data, but could also be an IP host provided with a MARP layer used to transmit IP packets over the network in the same way as made by IP host 10.

The MARP layer operates with a table called the MARP table represented in Fig. 2A. The MARP table maps the next hop IP address into a set of candidate IP addresses corresponding to candidate routers amongst the set of routers 20, 22 and 24 interfacing the IP network as illustrated in Fig. 1. In the simplest form, there is only one entry in the MARP table for the default router, that points on the set of candidate routers which can act as default routers.

The candidate routers associated with the IP addresses in MARP table can either be configured to the MARP layer via a configuration tool, or be dynamically acquired using a learning protocol such as an extension to the Dynamic Host configuration Protocol (DHCP).

As some ones of the candidate routers may not be active at a given time, the MARP layer uses the ARP table provided by the network layer as illustrated in Fig. 2B. The ARP table maps the IP addresses provided by the MARP table into network addresses.

Referring now to Fig. 3, the selection of an active router is as follows. When an IP packet is to transmit over the network, the MARP layer is called by the IP Layer and the next hop IP address (usually that of the default router) is provided as a parameter for looking up the MARP table (step 50). If the next hop IP address matches an entry in the MARP table (step 52), an associated list of candidate routers is built (step 54). The candidate routers are then checked in the ARP table, one by one (step 56). A determination is made (step 58) of the candidate routers which have a recent entry in the ARP table, and these routers are selected as active candidate routers. Note that, if no active candidate routers can be determined (step 58), the packet is destroyed (step 60).

Out of the list of the active candidate routers, the MARP layer selects (step 62) one IP address corresponding to a candidate router that is passed to the network layer as a substitute of the original next hop IP address as selected by the IP layer. In the preferred embodiment, this selection is performed on a per packet basis, without an history of previous selection, but this is not the only possible selection algorithm. Other techniques like round robin or byte wise weighting mechanisms could be used alternatively. The preferred implementation uses an hash coding technique as described in European Patent Application n° 98480062.3, in order to stick a TCP connection to a same candidate router as long as the candidate topology is left unchanged. The hash coding uses the

destination IP address and the pair of ports in the packets. These are mingled with the candidate routers IP addresses, one by one. The highest resulting hash value is selected. Weight coefficients may be used to modify the statistical expectancy of each individual candidate, in order to match their capacity.

At last, the IP packet is sent to the network layer (step 64) for it to be transmitted to the candidate router which has been selected. It must be noted that the IP packet will directly be sent to the network layer when no match has been found (step 52) in looking up the MARP table because the next hop IP address corresponds to a router or a host which is not required to be substituted.

It must be noted that the MARP layer only uses candidates that are already present in the ARP table. As a consequence, MARP layer uses an out-of-band technique to be sure that the ARP table is correctly filled with all the up-to-date information. In the preferred embodiment, periodic void packets like ICMP echo are transmitted to the non-active routers, that is candidate routers which are not present in the ARP table. Upon such packets, the ARP function in the network layer will automatically refresh the entry by using the ARP protocol. Also, at the initial time, one such packet is sent to all the configured routers to preset the ARP table before a single data is issued by an application layer.

The ARP function ensures the freshness of the ARP table by aging the entries and flushing the older ones. To maintain the status of active candidate routers, the preferred method consists in resetting the age of an entry each time a packet is received from a matching network address. Also, if an entry gets old, but before it is flushed by ARP, MARP may flush the ARP table entry right before it passes a packet to the Network layer with the next hop IP address pointing on that router. Again, this forces the Network layer to use ARP procedures to check for the router availability.

Claims

1. Data transmission system for transmitting packetized data from an IP host (10) having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein said IP host is connected to said IP network via a layer 2 network (18) interfacing said IP network by a set of routers (20, 22, 24) ;

   said system being characterized in that said IP host further includes a Multiple Address Resolution Protocol (MARP) layer (36) between said IP layer and said network layer for selecting one of said set of routers in response to the next hop IP address provided by said IP layer to said multiple ARP layer when a packet of data is to be transmitted from said IP host to one of said workstations.

2. Data transmission system according to claim 1, wherein said IP host (10) is provided with an Address Resolution Protocol (ARP) in charge of resolving any IP address into a network address of the router to be used in said layer 2 network (18) by mapping in an ARP table said IP address into the network address of an active router amongst said set of routers (20, 22, 24).

3. Data transmission system according to claim 2, wherein said MARP layer (36) includes a MARP table mapping said next hop IP address into a list of candidate routers amongst said set of routers (20, 22, 24), said candidate routers being mapped in said ARP table into active candidate routers able to be used as routers for transmitting said packet of data from said IP host (10) to one of said workstations (12, 14).

4. Data transmission system according to claim 3, wherein one router is selected amongst said active candidate routers by using a hash coding method based upon the destination IP address, the pair of source and destination

ports in said packet of data to be transmitted, and the active candidate router IP addresses.

5. Method of selecting a router by an IP host (10) in a data transmission system transmitting packetized data from said IP host having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein said IP host is connected to said IP network via a layer 2 network (18) interfacing said IP network by a set of routers (20, 22, 24) ;

said method being characterized by determining a list of candidate routers amongst said set of routers and determining a list of active candidate routers amongst said candidate routers before selecting said router to be used for transmitting said packet of data amongst said list of active candidate routers.

6. Method according to claim 5, wherein said step of determining said list of active candidate routers is performed by a Multiple Address Resolution Protocol (MARP) layer (36) between the IP layer (34) and the network layer of said IP host (10).

7. Method according to claim 6, wherein said step of determining said list of candidate routers is performed by said MARP layer (36) by a look up of a MARP table using the next hop IP address as entry.

8. Method according to claim 7, wherein said step of selecting said router to be used for transmitting said packet of data is performed by using a hash coding technique based upon the destination IP address, the pair of source and destination ports in said packet of data to be transmitted, and the active candidate router IP addresses.

Multiple ARP functionality for an IP data transmission system

Abstract

Data transmission system for transmitting packetized data from an IP host (10)having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein the IP host is connected to the IP network via a layer 2 network (18) interfacing the IP network by a set of routers (20, 22, 24). The IP host further includes a Multiple Address Resolution Protocol (MARP) layer (36) between the IP layer and the network layer for selecting one of the set of routers in response to the next hop IP address provided by the IP layer to the multiple ARP layer when a packet of data is to be transmitted from the IP host to one of the workstations.

Fig. 1

FR9-99-008
Lamberton et al
1/3

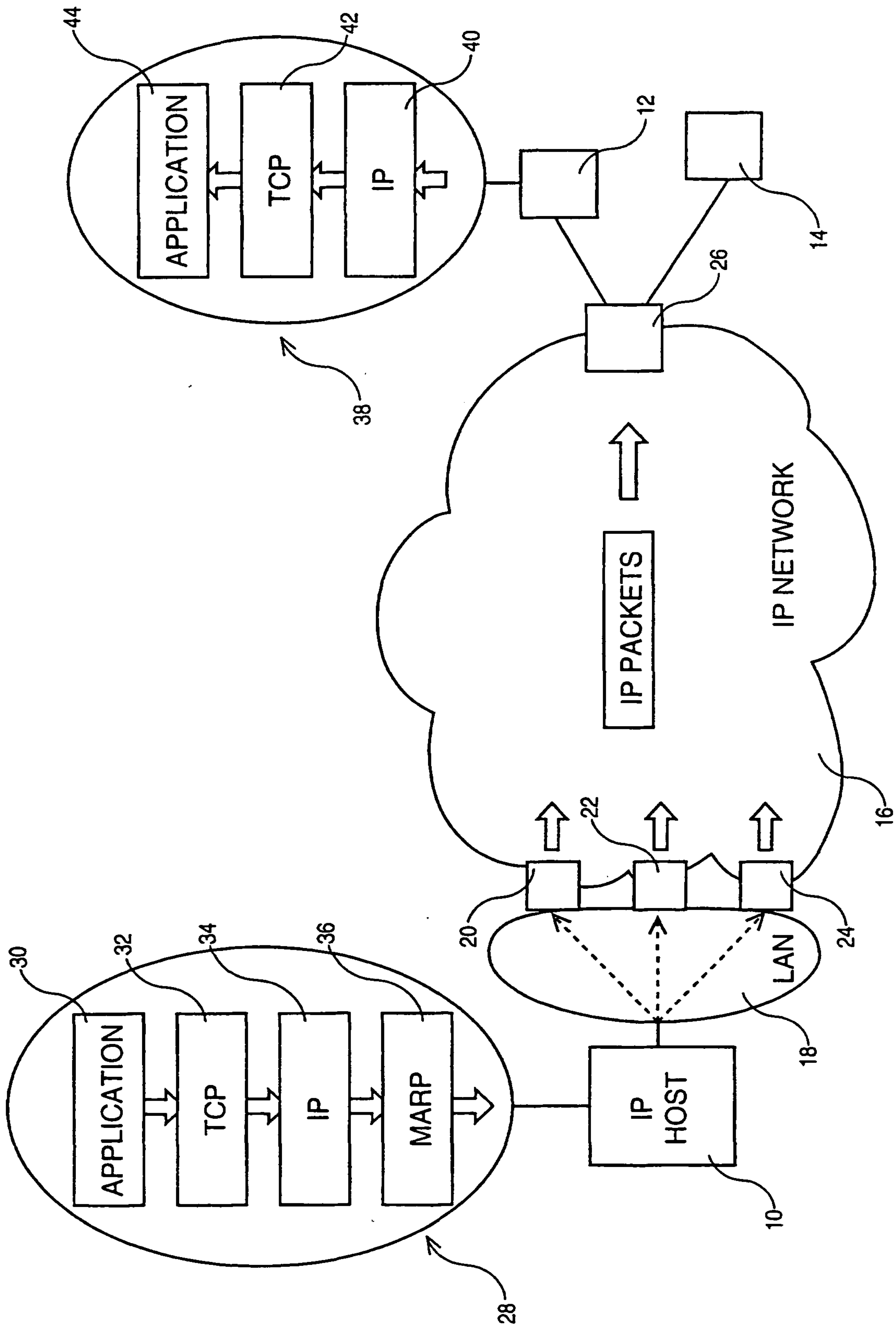

FIG. 1

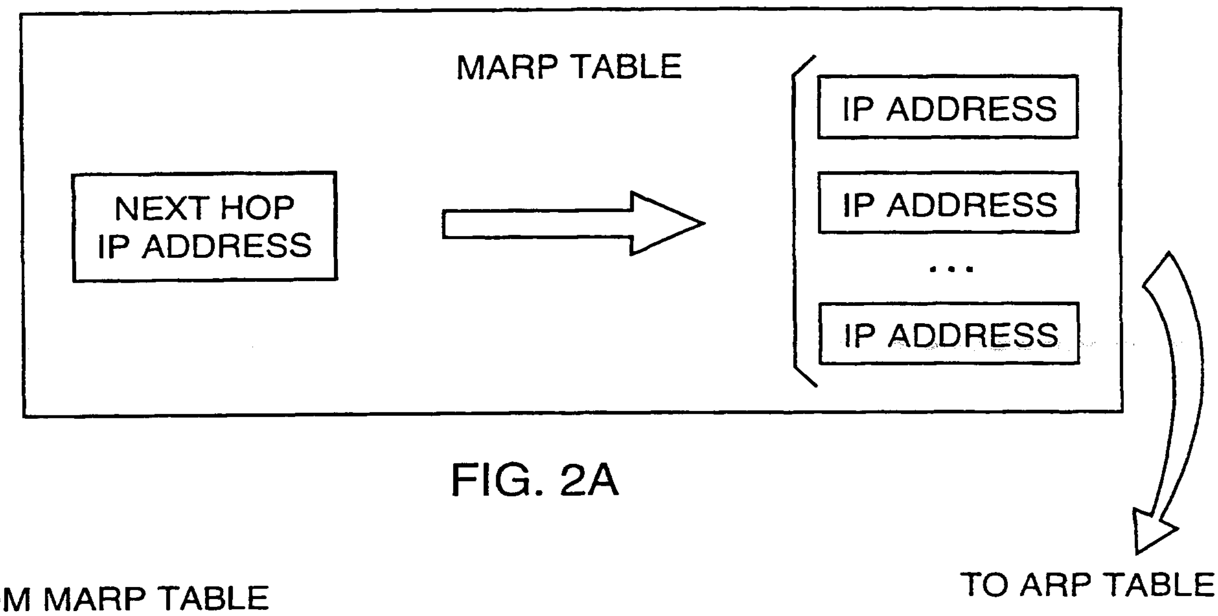
MARP TABLE
NEXT HOP IP ADDRESS
IP ADDRESS
IP ADDRESS
...
IP ADDRESS
TO ARP TABLE

FIG. 2A

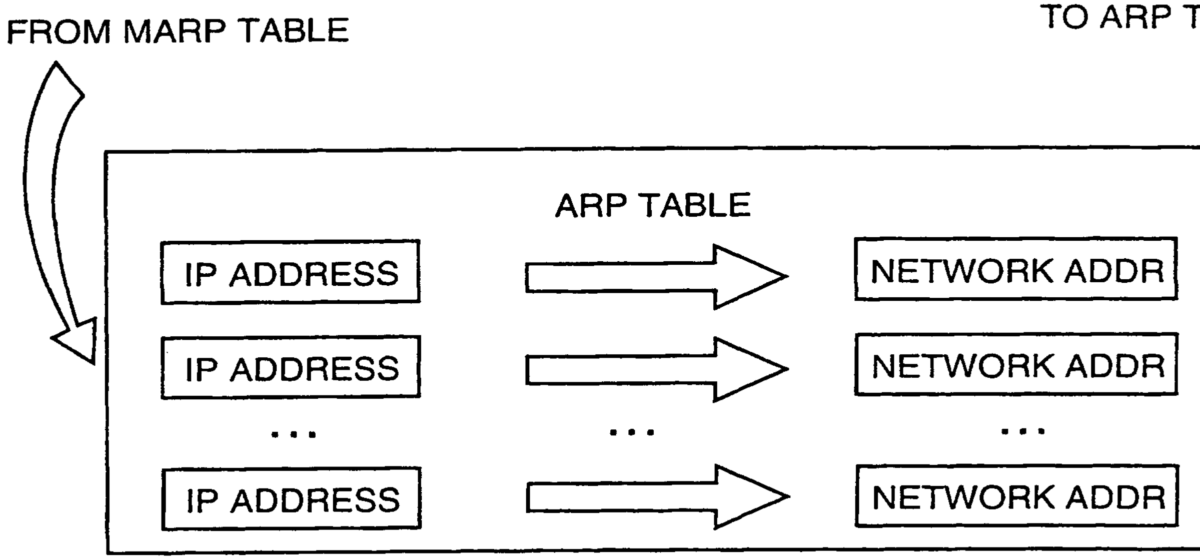
FROM MARP TABLE
ARP TABLE
IP ADDRESS
NETWORK ADDR
IP ADDRESS
NETWORK ADDR
...
...
...
IP ADDRESS
NETWORK ADDR

FIG. 2B

FR9-99-008
Lamberton et al
3/3

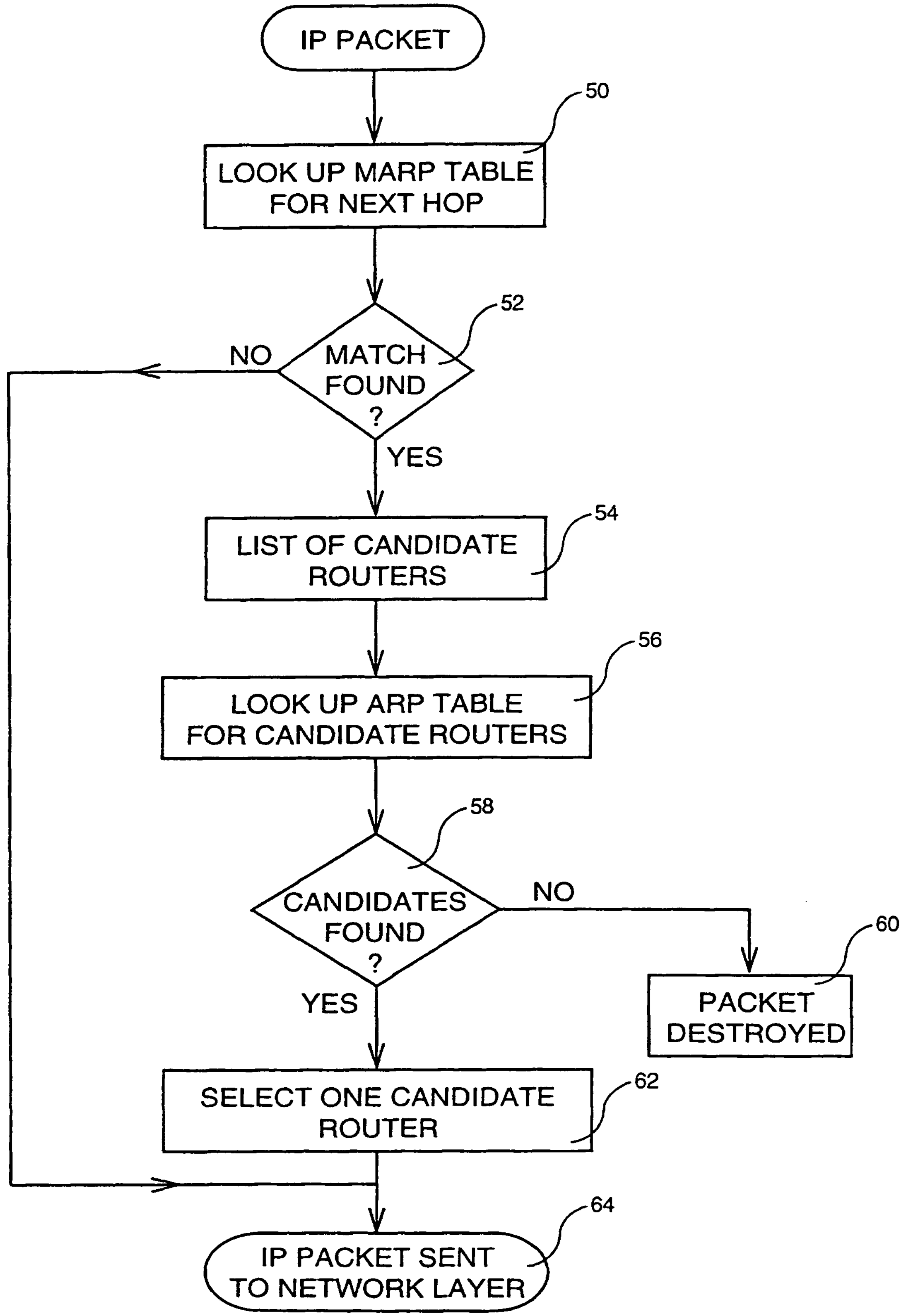

FIG. 3

EPO - DG 1

29. 04. 1999

IBM

Compagnie IBM France
Direction de la Propriété Intellectuelle
Le Plan-du-Bois
F 06610 La Gaude

REGISTERED LETTER + A.D.

(58)

La Gaude, 23 April 1999

EUROPEAN PATENT OFFICE
Directorate General I
B.P. 5818 Patentlaan, 2
2280 HV Rijswijk (Z.H.)
NETHERLANDS

INTERNATIONAL BUSINESS MACHINES CORPORATION

Patent Application N°	: **99 480 017.5**
Filing date	: 30 March 1999
Our Reference	: FR 9 99 008

Subject : **Payment fees**

Dear Sirs,

Please find enclosed for above referenced application a debit order corresponding to following payment of fees :

- filing
- search
- additional copy
- priority documents (2)

Very truly yours,

Yves Nicolas ETORRE
Authorized Representative

Attachments: debit order

md/0725

IBM FRANCE

INTELLECTUAL PROPERTY
CER - BP 13
06610 La Gaude

From : M. Domenech
Fax nbr : (33) 493 24 75 11

EPA EPO-OEB
DG 1
Reçu:
20-08-1999
11 | | ANL ZEICHN

Facsimile

To: Office Europeen des Brevets · La Haye
@Fax: (070) 3 40 30 16
From: Monique Domenech
Date: 20 Août 1999
Re: Copies officielles
Pages: 1, including this

Ref : 99480017.5 (FR999008)
99480015.9 (FR999018)

Messieurs,

Nous avons commandé des copies officielles pour lesdemandes de brevet ci-dessus référencées le 23 Avril 1999.
Il semblerait que nous n'ayons pas reçu ces copies à ce jour. Vous serait-il possible de nous faire savoir rapidement si celles-ci nous ont bien été envoyées.

Nous vous en remercions par avance.

Cordialement,

MDomenech

M. Domenech
Cie IBM France - Direction de la
Propriété Intellectuelle
06610 La Gaude - FRANCE
Tel : (33) 492.11.53.11
Fax : (33) 493.24.75.11

Multiple ARP functionality for an IP data transmission system

Technical field

The present invention deals with a new way for load balancing outgoing IP packets from an IP host such as a large Web server, and relates in particular to a multiple ARP functionality for an IP data transmission system.

Background

Modern digital networks are made to operate over different transmission media and interconnect upon request a very large number of users (e.g. hosts) and applications through fairly complex digital communication networks.

Due to the large variety of users' profiles and distributed applications, the traffic is becoming more and more bandwidth consuming, non-deterministic and requiring more connectivity. This has been the driver for the emergence of fast packet switching techniques in which data from different origins are chopped into fixed or variable length packets or datagrams, and then transferred, over high speed digital networks, between a data source and a target terminal equipment.

Several types of networks have been installed throughout the world, which need to be interconnected (e.g. via so called Routers) to optimize the possibilities of organizing traffic between source hosts and target hosts located anywhere in the world. This is made possible by using so-called internetworking.

Internetwork (also referred to as internet) facilities use a set of networking protocols such as Transmission Control Protocol/Internet Protocol (TCP/IP) developed to allow cooperating host computers to share resources across the internetwork. TCP/IP is a set of data communication protocols that are

referred to as internet protocol (IP) suite. Because TCP and IP are the best known, it has become common to use the term TCP/IP to refer to the whole protocol family. TCP and IP are two of the protocols in this suite. Other protocols of the suite are User Datagram Protocol (UDP), Address Resolution Protocol (ARP), Real Time Protocol (RTP) etc…

An internet may thus be a collection of heterogeneous and independent networks using TCP/IP, and connected together by routers. The administrative responsibilities for an internet (e.g. to assign IP addresses and domain names) can be within a single network (LAN) or distributed among multiple networks.

When a communication of data has to be established from a source host to a particular IP destination over an IP network, there is a number of methods to determine the first hop router of the network towards this destination. These include running (or snooping) dynamic routing protocol such as Routing Information Protocol (RIP) or Open Shortest Path First (OSPF) version, running an ICMP router discovery client or using a statically configured default route.

Running a dynamic routing protocol on every end-host may be infeasible for a number of reasons, including administrative overhead, processing overhead, security issues, or lack of a protocol implementation for some platforms. Neighbor or router discovery protocols may require active participation by all hosts on a network, leading to large timer values to reduce protocol overhead in face of large numbers of hosts. This can result in a significant delay in the detection of a lost (i.e., dead) neighbor, which may introduce unacceptably long "black hole" periods.

The use of a statically configured default route is quite popular, it minimizes configuration and processing overhead on the end-host and is supported by virtually every IP implementation. This mode of operation is likely to persist as Dynamic Host Configuration Protocols (DHCP) are deployed, which typically provide configuration for an end-host IP address and

default gateway. However, this creates a single point of failure. Loss of the default router results in a catastrophic event, isolating all end-hosts that are unable to detect any alternate path that may be available.

One solution to solve this problem is to allow hosts to appear to use a single router and to maintain connectivity even if the actual first hop router they are using fails. Multiple routers participate in this protocol and in concert create the illusion of a single virtual router. The protocol ensures that one and only one of the routers is forwarding packets on behalf of the virtual router. End hosts forward their packets to the virtual router. The router forwarding packets is known as the active router. A standby router is selected to replace the active router should it fail. The protocol provides a mechanism for determining active and standby routers, using the IP addresses on the participating routers. If an active router fails, a standby router can take over without a major interruption in the host's connectivity.

Another similar approach is the use of Virtual Router Redundancy Protocol (VRRP) designed to eliminate the single point of failure inherent in the static default routed environment. VRRP specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers on a LAN. The VRRP router controlling the IP address(es) associated with a virtual router is called the Master, and forwards packets sent to these IP addresses. The election process provides dynamic fail-over in the forwarding responsibility should the Master become unavailable. Any of the virtual router's IP addresses on a LAN can then be used as the default first hop router by end-hosts. The advantage gained from using VRRP is a higher availability default path without requiring configuration of dynamic routing or router discovery protocols on every end-host.

Unfortunately the two above solutions cannot provide load balancing for a given host's traffic because only the router that answered the ARP is used. Also, customers are reluctant

to change their main router configuration to enable such a function.

Summary of the invention

Accordingly, the object of the invention is to provide a data transmission system including an IP network wherein it is the IP host which selects directly the default router thereby improving load balancing and high availability.

Another object of the invention is to enable an IP source host to be aware of the availability of a set of candidate default routers and to select one of them dynamically, ensuring both load balancing and high availability.

Another object of the invention is a method of selecting a router amongst a set of routers for an IP host in a data transmission system including an IP network.

Therefore, the invention relates to a data transmission system for transmitting packetized data from an IP host having at least an IP layer and a network layer to a plurality of work-stations by the intermediary of an IP network and wherein the IP host is connected to the IP network via a layer 2 network interfacing the IP network by a set of routers, the IP host further including a Multiple Address Resolution Protocol (MARP) layer between the IP layer and the network layer for selecting one of the set of routers in response to the next hop IP address provided by the IP layer to the multiple ARP layer when a packet of data is to be transmitted from the IP host to one of the workstations.

Brief description of the drawings

The above and other objects, features and advantages of the invention will be better understood by reading the following more particular description of the invention in conjunction with the accompanying drawings wherein :

Fig. 1 represents schematically a data transmission system wherein an IP host can select one router amongst a set of routers according to the invention.

Fig. 2A and 2B represent respectively the MARP table and the ARP table used in combination to achieve the method according to the invention.

Fig. 3 is a flow chart of the method of selecting a router according to the invention.

Detailed description of the invention

In reference to Fig. 1, the invention is implemented in a data transmission system wherein an IP host has to transmit data to one or several workstations 12, 14 via an IP network 16 such as Internet. It can be assumed that IP host 10 is connected to IP network 16 by means of a layer 2 network such as Local Area Network (LAN) 18 which is interfacing IP network 16 by a set of input routers 20, 22 and 24. The IP packets are routed over the IP network via a plurality of routers (not shown) until an output router 26 connected directly (or by means of a layer 2 network) to workstations 12 or 14.

As illustrated in Fig. 1, to communicate over the IP network, IP host 10 must implement a layered set of protocols 28 referred as the Internet protocol suite. Without the invention the protocol suite would be used as follows :

- the application layer 30 (level 5) generates a data stream to be sent and passes this data stream to a transport layer,
- the transport layer (level 4) such as TCP layer 32, segments the data stream into packets and passes the packet to the IP layer for routing to the destination IP address with an added TCP Header,
- the IP layer 34 finds the next hop IP address based upon the destination IP address. Normally, with the IP Host

which does not run a routing protocol, this address is a default entry that leads to a default router.

- IP layer 34 passes the IP packet to the network layer (not shown) with an added IP header information. As a side parameter, the IP layer informs the network layer of the next hop IP address.
- the network layer resolves the next hop IP address into a network address of the default router using the ARP protocol and transmits the packet over the IP network.

The invention introduces a new layer between IP layer 34 and the network layer, a Multiple ARP (MRAP) layer 36. Therefore, IP layer 34 passes the packet and the next hop IP address to MARP layer 36 instead of the network layer. As explained below, this MARP layer runs an algorithm to determine the best physical router 20, 22 or 24 based on parameters defined in the packet such as source and destination addresses and ports.

At the destination workstation 12, a reciprocal protocol suite 38 is implemented. Namely, the network layer passes the IP packets to IP layer 40 which transfers the packets to TCP layer 42 for reassembling them into a data stream communicated to the application layer 44. Note that workstation 12 does not include a MARP layer since such a layer is not required for receiving data, but could also be an IP host provided with a MARP layer used to transmit IP packets over the network in the same way as made by IP host 10.

The MARP layer operates with a table called the MARP table represented in Fig. 2A. The MARP table maps the next hop IP address into a set of candidate IP addresses corresponding to candidate routers amongst the set of routers 20, 22 and 24 interfacing the IP network as illustrated in Fig. 1. In the simplest form, there is only one entry in the MARP table for the default router, that points on the set of candidate routers which can act as default routers.

The candidate routers associated with the IP addresses in MARP table can either be configured to the MARP layer via a configuration tool, or be dynamically acquired using a learning protocol such as an extension to the Dynamic Host configuration Protocol (DHCP).

As some ones of the candidate routers may not be active at a given time, the MARP layer uses the ARP table provided by the network layer as illustrated in Fig. 2B. The ARP table maps the IP addresses provided by the MARP table into network addresses.

Referring now to Fig. 3, the selection of an active router is as follows. When an IP packet is to transmit over the network, the MARP layer is called by the IP Layer and the next hop IP address (usually that of the default router) is provided as a parameter for looking up the MARP table (step 50). If the next hop IP address matches an entry in the MARP table (step 52), an associated list of candidate routers is built (step 54). The candidate routers are then checked in the ARP table, one by one (step 56). A determination is made (step 58) of the candidate routers which have a recent entry in the ARP table, and these routers are selected as active candidate routers. Note that, if no active candidate routers can be determined (step 58), the packet is destroyed (step 60).

Out of the list of the active candidate routers, the MARP layer selects (step 62) one IP address corresponding to a candidate router that is passed to the network layer as a substitute of the original next hop IP address as selected by the IP layer. In the preferred embodiment, this selection is performed on a per packet basis, without an history of previous selection, but this is not the only possible selection algorithm. Other techniques like round robin or byte wise weighting mechanisms could be used alternatively. The preferred implementation uses an hash coding technique as described in European Patent Application n° 98480062.3, in order to stick a TCP connection to a same candidate router as long as the candidate topology is left unchanged. The hash coding uses the

destination IP address and the pair of ports in the packets. These are mingled with the candidate routers IP addresses, one by one. The highest resulting hash value is selected. Weight coefficients may be used to modify the statistical expectancy of each individual candidate, in order to match their capacity.

At last, the IP packet is sent to the network layer (step 64) for it to be transmitted to the candidate router which has been selected. It must be noted that the IP packet will directly be sent to the network layer when no match has been found (step 52) in looking up the MARP table because the next hop IP address corresponds to a router or a host which is not required to be substituted.

It must be noted that the MARP layer only uses candidates that are already present in the ARP table. As a consequence, MARP layer uses an out-of-band technique to be sure that the ARP table is correctly filled with all the up-to-date information. In the preferred embodiment, periodic void packets like ICMP echo are transmitted to the non-active routers, that is candidate routers which are not present in the ARP table. Upon such packets, the ARP function in the network layer will automatically refresh the entry by using the ARP protocol. Also, at the initial time, one such packet is sent to all the configured routers to preset the ARP table before a single data is issued by an application layer.

The ARP function ensures the freshness of the ARP table by aging the entries and flushing the older ones. To maintain the status of active candidate routers, the preferred method consists in resetting the age of an entry each time a packet is received from a matching network address. Also, if an entry gets old, but before it is flushed by ARP, MARP may flush the ARP table entry right before it passes a packet to the Network layer with the next hop IP address pointing on that router. Again, this forces the Network layer to use ARP procedures to check for the router availability.

Claims

1. Data transmission system for transmitting packetized data from an IP host (10) having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein said IP host is connected to said IP network via a layer 2 network (18) interfacing said IP network by a set of routers (20, 22, 24) ;

   said system being characterized in that said IP host further includes a Multiple Address Resolution Protocol (MARP) layer (36) between said IP layer and said network layer for selecting one of said set of routers in response to the next hop IP address provided by said IP layer to said multiple ARP layer when a packet of data is to be transmitted from said IP host to one of said workstations.

2. Data transmission system according to claim 1, wherein said IP host (10) is provided with an Address Resolution Protocol (ARP) in charge of resolving any IP address into a network address of the router to be used in said layer 2 network (18) by mapping in an ARP table said IP address into the network address of an active router amongst said set of routers (20, 22, 24).

3. Data transmission system according to claim 2, wherein said MARP layer (36) includes a MARP table mapping said next hop IP address into a list of candidate routers amongst said set of routers (20, 22, 24), said candidate routers being mapped in said ARP table into active candidate routers able to be used as routers for transmitting said packet of data from said IP host (10) to one of said workstations (12, 14).

4. Data transmission system according to claim 3, wherein one router is selected amongst said active candidate routers by using a hash coding method based upon the destination IP address, the pair of source and destination

ports in said packet of data to be transmitted, and the active candidate router IP addresses.

5. Method of selecting a router by an IP host (10) in a data transmission system transmitting packetized data from said IP host having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein said IP host is connected to said IP network via a layer 2 network (18) interfacing said IP network by a set of routers (20, 22, 24) ;

said method being characterized by determining a list of candidate routers amongst said set of routers and determining a list of active candidate routers amongst said candidate routers before selecting said router to be used for transmitting said packet of data amongst said list of active candidate routers.

6. Method according to claim 5, wherein said step of determining said list of active candidate routers is performed by a Multiple Address Resolution Protocol (MARP) layer (36) between the IP layer (34) and the network layer of said IP host (10).

7. Method according to claim 6, wherein said step of determining said list of candidate routers is performed by said MARP layer (36) by a look up of a MARP table using the next hop IP address as entry.

8. Method according to claim 7, wherein said step of selecting said router to be used for transmitting said packet of data is performed by using a hash coding technique based upon the destination IP address, the pair of source and destination ports in said packet of data to be transmitted, and the active candidate router IP addresses.

Multiple ARP functionality for an IP data transmission system

Abstract

Data transmission system for transmitting packetized data from an IP host (10)having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein the IP host is connected to the IP network via a layer 2 network (18) interfacing the IP network by a set of routers (20, 22, 24). The IP host further includes a Multiple Address Resolution Protocol (MARP) layer (36) between the IP layer and the network layer for selecting one of the set of routers in response to the next hop IP address provided by the IP layer to the multiple ARP layer when a packet of data is to be transmitted from the IP host to one of the workstations.

Fig. 1

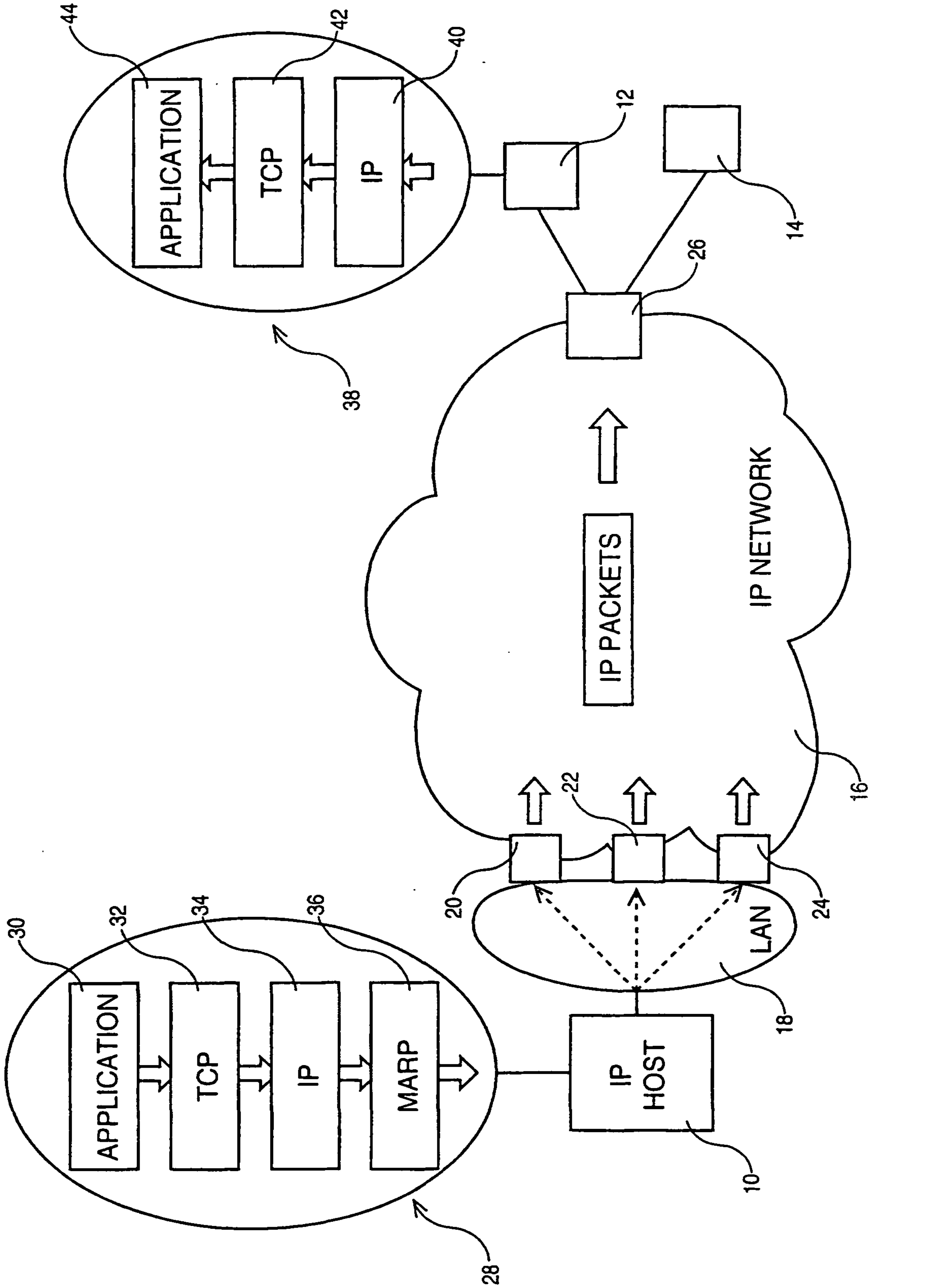
APPLICATION
TCP
IP
44
42
40
38
12
14
26
IP PACKETS
IP NETWORK
16
22
20
24
LAN
18
APPLICATION
TCP
IP
MARP
30
32
34
36
28
IP HOST
10

FIG. 1

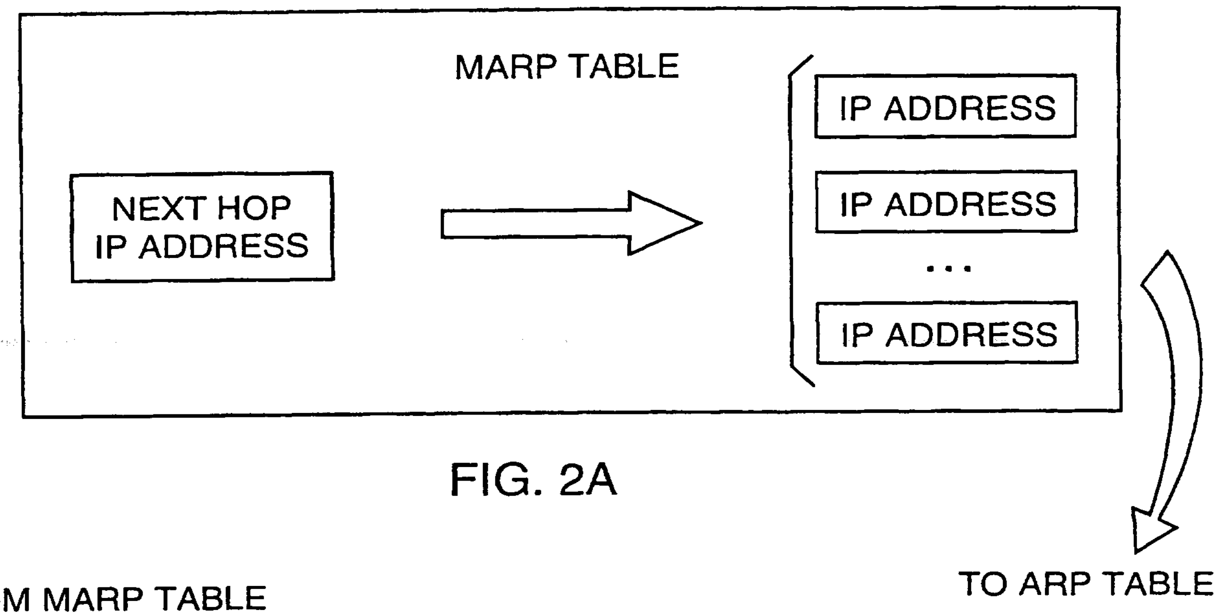
MARP TABLE
NEXT HOP
IP ADDRESS
IP ADDRESS
IP ADDRESS
...
IP ADDRESS
TO ARP TABLE

FIG. 2A

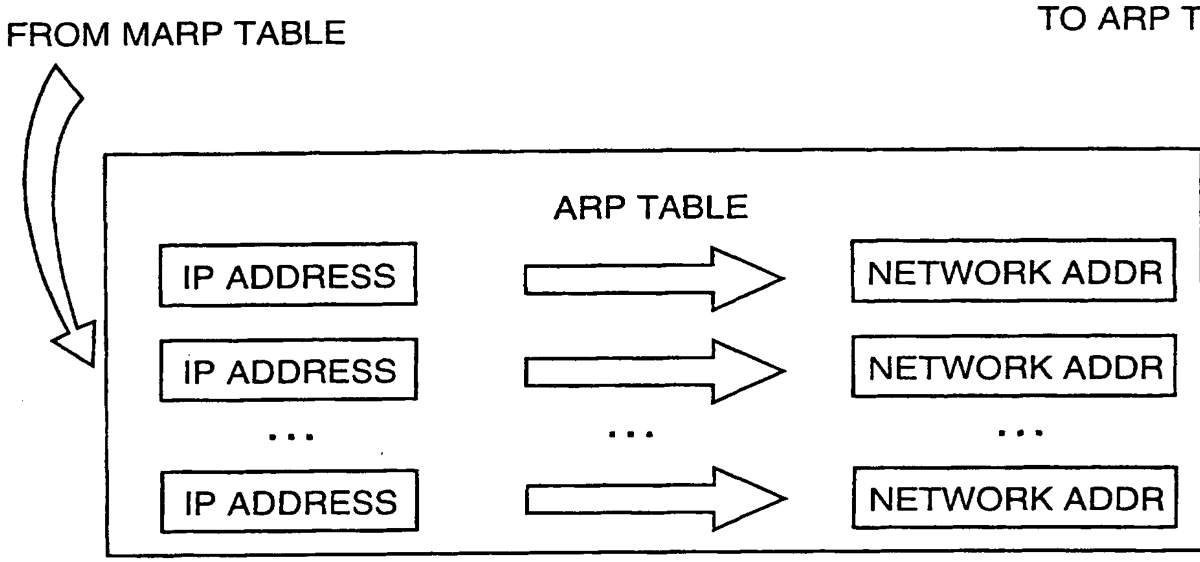
FROM MARP TABLE
ARP TABLE
IP ADDRESS
NETWORK ADDR
IP ADDRESS
NETWORK ADDR
...
...
...
IP ADDRESS
NETWORK ADDR

FIG. 2B

FR9-99-008
Lamberton et al
3/3

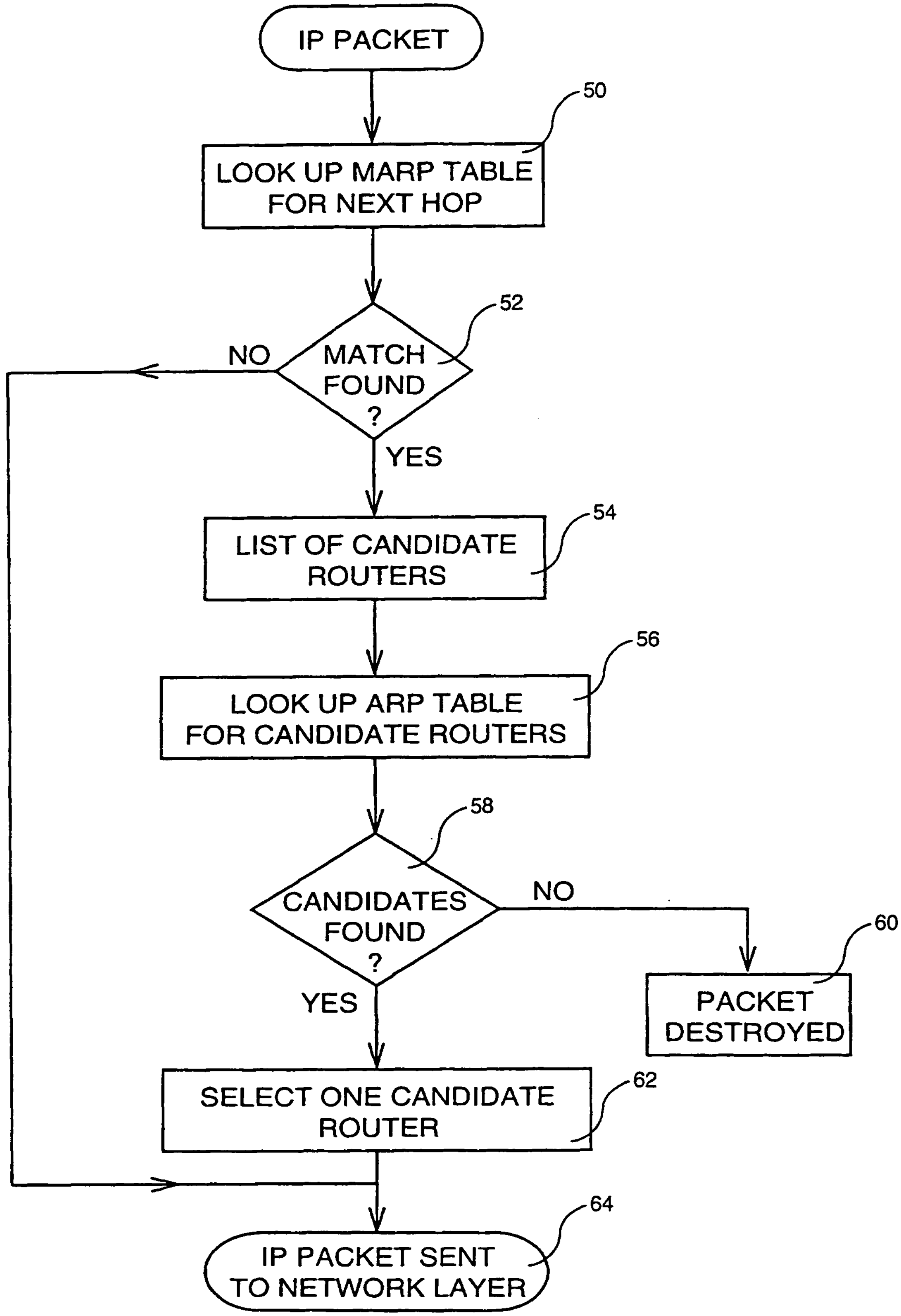

FIG. 3